

In this work I search through interviews of Czech witnesses of the holocaust from the MALACH project to find relevant parts of these testimonies. Audio records of these interviews are automatically recognized by a system for an automatic speech recognition. Automatically recognized texts are then lemmatized and tagged. In this work I present a script which generates parametrizable collections of documents from these preprocessed texts. The task of unsegmented speech retrieval is then reformulated to a task of information retrieval in this collections of documents. In this work, I describe many experiments which examine the influence of different retrieval techniques on retrieval results on this data collection. Mainly, I study an influence of a morphological normalization (lemmatization), different types of IR systems (TF-IDF model, Okapi model and Indri model), blind relevance feedback, stopword list based on frequencies of terms and part-of-speech categories. I also place emphasis on various values of length and overlap parameters of generated documents. The results of these experiments are verified on test data. Audio records, outputs from automatic speech recognition system and topics for information retrieval are not part of this work due to legal grounds.